REMARKS

I. Status of the Claims

Claims 4, 6-8, 17, 20-28, and 48 are pending in the application. A Request for Continuing Examination accompanies this reply. Applicants thank Examiner Crouch and Examiner Lieto for taking time to discuss the outstanding issues with applicants' representatives.

In a final Office Action dated July 28, 2005, claims 4, 6-8, 17, 20-28 and 48 are rejected under 35 U.S.C. § 112, first paragraph as failing to comply with the enablement requirement. Claim 48 is objected to as being dependent upon a cancelled base claim. Applicants amended claim 48 in a Reply dated October 27, 2005.

II. Claim Objection

Claim 48 was objected to as being dependent upon a cancelled base claim. In applicants' Reply dated October 27, 2005, claim 48 was amended to depend from claim 6. In an Advisory Action dated November 30, 2005, the examiner entered the amendment and withdrew the objection in view of applicants' amendment.

III. The Claims are Patentable under 35 U.S.C. § 112, Paragraph 1

Claims 4, 6-8, 17, 20-28 and 48 were rejected under 35 U.S.C. § 112, first paragraph, for allegedly containing non-enabled subject matter. The examiner alleges that the specification, while being enabling for a recombinant nucleic acid molecule consisting of a nucleotide sequence encoding hepatitis C virus nonstructural proteins NS3, NS4 and NS5, wherein said nucleotide sequence is operably linked to regulatory elements, said regulatory elements comprising a promoter, enhancer, polyadenylation sequence, and at most the 9 most 3' nucleotides of the 5'-UTR of a hepatitis C virus, does not reasonably provide enablement for a recombinant nucleic acid molecule consisting of a nucleotide sequence encoding hepatitis C virus nonstructural proteins NS3, NS4 and NS5, wherein said nucleotide sequence is operably linked to regulatory elements, said regulatory elements comprising a promoter, enhancer, polyadenylation sequence, and a 5' untranslated region from any gene. Applicants traverse the rejection.

Applicants' claimed invention is a recombinant nucleic acid molecule consisting of a nucleotide sequence encoding hepatitis C virus nonstructural proteins NS3, NS4 and NS5, wherein said nucleotide sequence is operably linked to regulatory elements, said regulatory elements comprising a promoter, enhancer, polyadenylation sequence, and a hepatitis C virus 5'-untranslated region (5'-UTR). Furthermore, the claimed invention is a method of inducing an immune response against hepatitis C virus in a human uninfected by hepatitis C virus comprising administering the recombinant nucleic acid molecule. Contrary to the examiner's rejection, the specification does enable a person skilled in the art to practice the claimed invention, in part, wherein the regulatory elements comprise a hepatitis C virus 5'-UTR. The specification is enabling for regulatory elements comprising more than the "9 most 3' nucleotides of the 5'-UTR of a hepatitis C virus." The specification states that the 5'-UTR can include the last 9 nucleotides of the HCV 5'-UTR, the last 50 nucleotides, the last 100 nucleotides, the last 150 nucleotides, the last 200 nucleotides the last 250 nucleotides, the last 300 nucleotides, or the entire HCV 5'-UTR. See specification, for example, page 10, line 18 to page 11, line 2. These proportionate lengths of HCV 5'-UTR are functional in an expression construct comprising a nucleic acid encoding the HCV NS3, NS4, and NS5 genes. Looking to the specification as filed and to publications as of the filing date of the application, one of skill in the art would understand that expression constructs incorporating the HCV 5'-UTR have transcriptional activity.

The amount of guidance or direction needed to enable the invention is inversely related to the amount of knowledge in the state of the art as well as the predictability in the art. *In re Fisher* 427 F.2d 833,839, 166 USPQ 18,24 (CCPA 1970). MPEP 2164.03. Given the state of the art at the time of filing the application, one would be able to use the HCV 5'-UTR in a recombinant expression construct comprising a nucleic acid encoding the HCV NS3, NS4, and NS5 genes. Yoo et al., *Virology* **191:** 889-899, 1992 (Exhibit A; of record in a Supplemental Information Disclosure Statement filed October 27, 2005) further supports the enabling disclosure of the specification. Following the teaching of the Yoo et al. reference, one would have a reasonable expectation of success in constructing a recombinant nucleic acid molecule including HCV NS3, NS4, and NS5 genes, and regulatory elements including portions of the HCV 5'-UTR, as claimed. The Yoo et al. reference measured transient expression from an expression construct with a promoter region including various regions of the HCV 5'-UTR

and a chloramphenicol transferase (CAT) reporter. These experiments mapped the *cis*-acting elements controlling translation in the HCV genome linking a full length (nucleotides 1 to 341) or deleted versions of the 5'-UTR of HCV RNA to the coding region of CAT mRNA. See, for example, Figure 1 of Yoo et al. The nucleotide sequence of the 5'-UTR of HCV RNA corresponds to SEQ ID NO: 2 (nucleotides 1 to 341) in the specification. The Yoo et al. reference identifies "an efficient positive control element that stimulates translation may be present downstream from nucleotide 255. This 86-nucleotide sequence contains a 28-nucleotide sequence at position 291 to 281 with 90% sequence identity... to a PEST-IV [element]." See for example, Yoo et al., paragraph spanning pages 893 to 894 and Figure 3. Therefore, a skilled practitioner would predict that a construct containing regions of the 5'-UTR of HCV, *e.g.*, the PEST-IV element, would produce increased levels of HCV NS3, NS4, or NS5 protein compared to a construct lacking the 5'-UTR of HCV and thus increase the likelihood of producing a protective immune response in a mammal.

As further stated in the Declaration by Dr. Jack Wands in Paper No. 13:

> "Utilizing information provided in the subject application and in Yoo et al. *Virology* **191**: 889-899, 1992, one skilled in the art would understand the function of the 5'-UTR of hepatitis C virus, including the positive and negative translational control elements within the 5'-UTR. One skilled in the art would be able to operably link the 5'-UTR of hepatitis C virus to a recombinant nucleic acid molecule acting as an expression plasmid for proteins, for example, hepatitis C virus non-structural (NS) protein."

Declaration of Dr. Jack Wands under 37 C.F.R. § 1.132 in Paper No. 13.

A person of skill in the art at the time of filing the application would be able to utilize an HCV 5'-UTR from various strains or isolates of hepatitis C virus as a regulatory element in a recombinant nucleic acid molecule, as claimed. The HCV 5'-UTR is sufficiently conserved or analogous from various HCV strains or isolates that one of skill in the art would be able to operably link the 5'-UTR of hepatitis C virus to a recombinant nucleic acid molecule acting as an expression plasmid for proteins, for example, hepatitis C virus non-structural (NS) protein. Bukh et al., *Proc. Natl. Acad. Sci. USA* **89:** 4942-4946, 1992 (Exhibit B; of record in an Information Disclosure Statement filed February 5, 2001) provides an alignment of nucleotide sequences of the 5'-UTR of 44 HCV isolates from around the world and further compares these 44 sequences to an additional 37 published sequences of 5'-UTR

from HCV isolates. See for example, Figure 1 and Figure 3 of the Bukh et al. reference. The sequences share three highly conserved domains and conserved open reading frames within the 5'-UTR of the HCV isolates. See Abstract and Figure 2. These HCV isolates share sequence similarities and are classified within various HCV groups, for example, HCV group I, HCV group II, or HCV group III. See Bukh et al. reference, page 4944, top left column. The authors hypothesize that the conserved ORFs are maintained because of a role in control of translation. See Bukh et al. reference, page 4944, right column. The Yoo et al. reference confirms this and states that a positive control element is present downstream of nucleotide 255, containing a PEST-IV element. In Figure 1 of the Bukh et al. reference, a sequence comparison of the 44 isolates shows that a 100 nucleotide downstream region (-100 to -1) is highly conserved among the 44 isolates studied in the Bukh et al. reference and likely contains the same positive control element, PEST-IV, as described in the Yoo et al. reference. Therefore, the Bukh et al. reference and the Yoo et al. reference provided sufficient evidence to enable the use of any HCV 5'-UTR as a regulatory element in the recombinant nucleic acid, as claimed. The HCV 5'-UTR is sufficiently conserved or analogous such that one of skill in the art would know which regions of the 5'-UTR to operably link as a regulatory element in the recombinant nucleic acid.

The examiner argues that multiple premature start codons present in the 5'-UTR of the HCV would have a negative effect on the efficiency of translation. Utilizing information provided in the subject application and in the Yoo et al. reference, one skilled in the art would understand the negative effect on translation in the region of ORF1, ORF2, ORF3, and ORF4 within the 5'-UTR of HCV and the positive effect on translation in the region of PESTI-IV within the 5'-UTR of HCV. See for example, Yoo et al. reference, Figure 1. Contrary to the examiner's assertion, one skilled in the art would be able to operably link the 5'-UTR of hepatitis C virus to a recombinant nucleic acid molecule acting as an expression plasmid for proteins, for example, hepatitis C virus non-structural (NS) protein.

Therefore a person of skill in the art would know how to utilize the HCV 5'-UTR to enhance expression of a recombinant nucleic acid encoding the HCV NS3, NS4, and NS5 proteins. The specification enables one of skill in the art to utilize various regions of the 5'-UTR of a hepatitis C virus (*i.e.*, more than "at most the 9 most 3' nucleotides of the 5'-UTR of a hepatitis C virus") to construct a recombinant nucleic acid molecule as claimed.

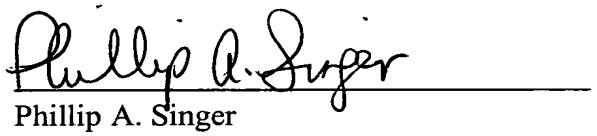
Furthermore, the specification enables one of skill in the art to utilize various regions of the 5'-UTR of a hepatitis C virus to construct and use the recombinant nucleic acid molecule encoding HCV NS3, NS4, and NS5 in a method of inducing an immune response against hepatitis C virus in a human uninfected by hepatitis C virus. Accordingly, applicants respectfully request that the rejection of claims 4, 6-8, 17, and 20-28 and 48 under 35 U.S.C. § 112, first paragraph, be withdrawn.

IV. Conclusion

In view of the foregoing, Applicants believe all claims now pending in this Application are in condition for allowance. The issuance of a formal Notice of Allowance at an early date is respectfully requested.

If the examiner believes a telephone conference would expedite prosecution of this application, please telephone the undersigned at 206-332-1380.

Date: January 27, 2006



Phillip A. Singer
Registration No. 40,176

Woodcock Washburn LLP
One Liberty Place - 46th Floor
Philadelphia PA 19103
Telephone: (215) 568-3100
Facsimile: (215) 568-3439

VIROLOGY 191, 889–899 (1992)

5′ End-Dependent Translation Initiation of Hepatitis C Viral RNA and the Presence of Putative Positive and Negative Translational Control Elements within the 5′ Untranslated Region

BYOUNG J. YOO,*[,1] RICHARD R. SPAETE,*[,2] ADAM P. GEBALLE,†
MARK SELBY,* MICHAEL HOUGHTON,* AND JANG H. HAN*[,3]

*Chiron Corporation, 4560 Horton Street, Emeryville, California 94608; and †Department of Molecular Medicine, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington 98104

Received July 23, 1992; accepted August 26, 1992

Hepatitis C virus (HCV) is a distant relative of pestiviruses and flaviviruses, but it has a 5′ untranslated region (UTR) with some features structurally similar to that of picornaviruses. In order to test the role of the 5′ UTR in controlling the expression of the HCV polyprotein, we fused full-length or deleted versions of the 5′ UTR of HCV-1 RNA to chloramphenicol acetyl transferase (CAT) mRNA to monitor CAT activity *in vivo*. We found: (1) the full-length 5′ UTR of HCV-1 RNA is translationally inactive while 5′ deletions which mimic a 5′ subgenomic RNA detected *in vivo* are active, (2) an efficient *cis*-acting element which represses translation is found at the 5′ terminus, (3) a putative element which enhances translation is found near the 3′ terminus of the 5′ UTR, (4) additional *cis*-acting elements including small open reading frames (ORFs) upstream from the putative enhancer element downregulate translation. We did not find evidence supporting the existence of an internal ribosome entry site in the 5′ UTR of HCV-1 RNA. These data suggest that HCV may employ a distinctive translation control strategy such as the generation of subgenomic viral mRNA in infected cells. Translational control of HCV might be responsible for some of the characteristic pathobiology seen in viral infection.

INTRODUCTION

HCV is the major etiologic agent of non-A, non-B hepatitis worldwide (Alter *et al.*, 1989; Choo *et al.*, 1989, 1990; Kuo *et al.*, 1989). It has a positive-strand RNA genome of approximately 9500 nucleotides which encodes a polyprotein that is processed into structural and nonstructural proteins (Choo *et al.*, 1989, 1990, 1991). HCV resembles flaviviruses and pestiviruses in genome organization within the polyprotein region and is proposed to be a member of the *Flaviviridae* (Choo *et al.*, 1990, 1991; Han *et al.*, 1991; Houghton *et al.*, 1991). HCV isolates show considerable amino acid sequence variations in coding regions, but they can be segregated into at least four different groups based on distinct amino acid sequence patterns (Houghton *et al.*, 1991; Okamoto *et al.*, 1992).

The 5′ untranslated region (UTR) of full-length HCV RNA appears to be 341 nucleotides long, based on at least five putative full-length HCV clones reported to date (Chen *et al.*, 1992; Han *et al.*, 1991; Tanaka *et al.*, 1992; Okamoto *et al.*, 1991). Unlike the polyprotein region, the 5′ UTR of HCV isolates are highly conserved (>98% within a group or >93% between groups), suggesting a functional importance (Han *et al.*, 1992). This region contains up to five upstream open reading frames (ORFs), the first four of which are overlapping in HCV-1, the prototype HCV isolate (Choo *et al.*, 1991; Han *et al.*, 1991). The 5′ UTR is homologous in nucleotide sequence to pestiviruses, especially in four regions (PEST-I to -IV) (Fig. 1C) (Han *et al.*, 1991). Previously, primer extension analysis has revealed that two prominent species of HCV RNA exist in samples derived from infected patients (Han *et al.*, 1991): a longer presumptive full-length genomic RNA, the 5′ terminus of which is predicted to form a hairpin structure (Chen *et al.*, 1992; Han *et al.*, 1991; Inchauspe *et al.*, 1991; Okamoto *et al.*, 1991, 1992), and a shorter 5′ subgenomic RNA, the 5′ terminus of which starts 145 nucleotide from the 5′ terminus of the longer RNA (Han *et al.*, 1991). These features suggest that control element(s) important to viral replication and polyprotein translation may be present in this region of the HCV genome.

The 5′ end of the HCV genome is considerably different from flaviviruses (Han *et al.*, 1991). However, it has been suggested that the presence of small upstream ORFs and the relatively long length of the HCV leader (Choo *et al.*, 1991; Han *et al.*, 1991; Inchauspe *et al.*,

[1] B.J.Y. is on sabbatical leave from the Department of Biology, Taegu University, Taegu, Korea.

[2] Present address: Vector Pharmaceuticals, Inc., 1815 Old County Road, Belmont, California 94002.

[3] To whom reprint requests should be addressed.

0042-6822/92 $5.00

1991; Kato *et al.*, 1990; Takamizawa *et al.*, 1991) are features reminiscent of poliovirus 5′ UTRs (Kitamura *et al.*, 1981). The 5′ UTRs of picornaviruses allow for cap-independent translation of picornaviral genomes and also contain a *cis*-acting site which allows for internal entry of ribosomes (Jang *et al.*, 1989; Pelletier and Sonenberg, 1988) in contrast to the more usual scanning mechanism hypothesized to account for the translation of most eukaryotic capped messages (Kozak, 1983, 1989). In fact, Tsukiyama-Kohara *et al.* (1992) recently reported the presence of an internal ribosome entry site (IRES) within the 5′ UTR of HCV RNA prepared from two Japanese isolates using an *in vitro* system. However, we report evidence using an *in vivo* system employing both monocistronic and dicistronic contructs, which indicates that the translation of HCV-1 RNA is mediated not by an internal initiation but by a 5′ end-dependent initiation which favors the conventional cap-dependent ribosome scanning mechanism. Moreover, we have mapped three distinct *cis*-acting control elements, which can block, repress, or enhance translation. Our results suggest that a subgenomic mRNA which is generated by an unknown mechanism *in vivo* may be the major template for the translation of the viral polyprotein.

MATERIALS AND METHODS

Cells, bacterial strains, and plasmids

Huh7, HeLa, and HepG2 cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% bovine calf serum (GIBCO-BRL, Gaithersburg, MD). Cells were grown in the presence of 7% CO_2. All plasmids were grown in *Escherichia coli* HB101, purchased from GIBCO-BRL.

Enzymes

Restriction enzymes and T4 DNA ligase were purchased from Boehringer-Mannheim (Indianapolis, IN), Taq-polymerase from Perkin Elmer (Norwalk, CT), and T7 RNA polymerase and RNasin from Promega (Madison, WI).

Construction of expression plasmids

The construction of plasmid pT7EMCAT and pSV_2CAT have been described (Elroy-Stein *et al.*, 1989; Gorman *et al.*, 1982). Plasmid pHCVCAT was constructed by attaching *Hin*dIII sites at the both ends of the 5′ UTR of HCV cDNA (Han *et al.*, 1991) by PCR (Saiki *et al.*, 1988) and cloning the resultant fragment into the *Hin*dIII site of pSV_2CAT. Plasmid pEQ355 was constructed by inserting the 341 bp 5′ UTR of HCV-1 into the *Hin*dIII/*Asp*718 sites resident in the multiple cloning site of β-galactosidase (β-gal) expression plasmid, pEQ176 (Schleiss *et al.*, 1991). The HCV-1 5′ UTR was generated as a *Hin*dIII/*Asp*718 PCR fragment using β114, an *Eco*RI fragment from a lambda vector (Houghton and Lee, unpublished data), as template. Plasmid pEQ391 [pCMV(CAT/HCV/LacZ)], was generated by ligating a 716-bp *Hin*dIII/*Ban*I fragment encoding the CAT gene isolated from plasmid pSV_2CAT (Gorman *et al.*, 1982) into plasmid pEQ355 at the *Hin*dIII site. The *Hin*dIII sites were ligated and the *Ban*I site and unligated *Hin*dIII site in pEQ355 were blunted with Klenow and religated. Plasmid pEQ416 [pCMV(CAT/polio/LacZ)] was constructed by ligating a 716-bp *Hin*dIII/*Bam*HI CAT-gene-encoding PCR fragment generated using pSV_2CAT as template, a 995-bp *Bam*HI/*Xho*I fragment encoding the poliovirus 5′ UTR isolated from pEQ396 (Spaete *et al.*, unpublished data), along with β-gal expression plasmid pEQ176 digested at the *Hin*dIII/*Xho*I site in the polylinker. pEQ396 is a β-gal expression plasmid constructed by cloning the 5′ UTR poliovirus sequence taken from pLNPOZ (Adam *et al.*, 1991) as an *Xho*I/*Pst*I fragment blunted using Klenow into pEQ377 digested at *Xba*I/*Sna*BI sites in the polylinker. The *Xba*I site was also filled with Klenow to create blunt ends. Transcription of β-gal in pEQ377 is promoted by T7 bacteriophage promoter (Geballe, unpublished data). Plasmid p(CAT/SV40/LacZ) was constructed by ligating the 716-bp *Hin*dIII/*Bam*HI CAT gene encoding PCR fragment described above, along with SV40 polyadenylation signals contained in an 847-bp *Bgl*II/*Bam*HI fragment isolated from pPR25 (Burke, unpublished data), and β-gal expression plasmid pEQ176 digested with *Hin*dIII/*Bgl*II. The authenticity of all PCR products was verified by sequencing each of the resulting segments in the plasmids (Chen and Seeburg, 1985).

Construction of hybrid CAT RNAs

Segments of pSV_2CAT vectors indicated by arrows (Fig. 1) were amplified by PCR as described (Saiki *et al.*, 1988; Shyamala and Ames, 1991). Each sense primer (PSV or P1 to P9) was designed to have a bacteriophage T7 promoter (TAATACGACTCACTATAG) at the 5′ end and a SV40 or HCV sequence of 16 to 18 bases at the 3′ end (Table 1). An antisense primer (PT30, Table 1) had a stretch of 40 Ts at the 5′ end and a complementary SV40 sequence (GGAGGAGTAG) at the 3′ end; this sequence binds to vectors 350 bp after a stop codon in the CAT gene by virtue of a perfect match of 10 nucleotides and an additional poly A track present in the template DNA. A segment of pT7EMCAT was amplified by primers T7 and T30 (Table 1). Each PCR product was transcribed by T7 polymerase with or

TABLE 1

LIST OF OLIGONUCLEOTIDES USED AS PRIMERS AND PROBE

Oligonucleotide	Size (nucleotide)	Position in HCV genome	Sequence (5′ to 3′)
P1	34	1 to 16	TAATACGACTCACTATAGGCCAGCCCCCTGATGG
P2	36	23 to 40	TAATACGACTCACTATAGCACTCCACCATGAATCAC
P3	36	35 to 52	TAATACGACTCACTATAGAATCACTCCCCTGTGAGG
P4	36	83 to 100	TAATACGACTCACTATAGCCATGGCGTTAGTATGAG
P5	36	99 to 116	TAATACGACTCACTATAGAGTGTCGTGCAGCCTCCA
P6	37	145 to 162	TAATACGACTCACTATAGGGTCTGCGGAACCGGGTA
P7	36	218 to 235	TAATACGACTCACTATAGCCTGGAGATTTGGGCGTG
P8	36	255 to 272	TAATACGACTCACTATAGGAGTAGTGTTGGGTCGCG
P9	36	322 to 339	TAATACGACTCACTATAGGGTCTCGTAGACCGTGCA
PSV	36	—	TAATACGACTCACTATAGATTCCAGAAGTAGTGAGG
T7	18	—	TAATACGACTCACTATAG
PT30	50	—	T_{40}GGAGGAGTAG
T30	45	—	T_{30}CAGGCGTAGCACCAG
JHC271	30	—	GGGATATATCAACGGTGGTATATCCAGTGA

Note. The position of each oligonucleotide in the HCV genome is based on Han *et al.* (1991).

without cap analogue (Promega, p2010), treated with DNase, extracted with phenol–chloroform, and precipitated twice with ethanol in the presence of 2.5 *M* ammonium acetate. Concentration of each poly(A)+ RNA was estimated by uv absorption and confirmed by Northern and dot–blot hybridization (Fig. 2) as described (Han *et al.*, 1986) using JHC271 as a probe (Table 1). In SV*CAT, R11, R13, and R14, sequences were internally inserted or deleted by an overlapping PCR method (Shyamala and Ames, 1991). The PCR products were confirmed to be correct by sequencing.

Translation of hybrid CAT RNAs *in vitro*

Synthetic RNAs were translated in nuclease-treated rabbit reticulocyte lysate (GIBCO-BRL) in the presence of 140 m*M* potassium acetate, as suggested by the manufacturer. Additional studies examining the influence of K^+ ion concentration on cap dependence were done in the presence of 50, 100, 150, and 200 m*M* potassium acetate. Aliquots of the translation product labeled with [^{35}S]methionine were analyzed by electrophoresis in a 12% polyacrylamide gel as previously described (Laemmli, 1970).

Transfection of hybrid CAT RNAs into mammalian cells for CAT assay

Two micrograms of each synthetic RNA was transfected into 1 × 10^6 cells in a 3.5-cm Costar plate (Thomas Scientific, Swedesboro, NJ) using 15 μg of lipofectin (GIBCO-BRL) according to the procedure of Felgner *et al.* (1987) modified by the manufacturer. Cells were incubated overnight and harvested for CAT assay as previously described (Gorman *et al.*, 1982). The relative CAT activity was shown to be linear between 0.5 and 5 μg of transfected RNA. Post-transfection incubation between 6 hr and overnight did not significantly affect CAT activity. For translation of RNAs in poliovirus-infected cells, Huh7 cells were infected with poliovirus (Mahoney strain, ATCC VR-59) at a multiplicity of infection (m.o.i.) of 100. Cells were transfected with RNAs 2 hr after the infection and harvested 4 hr after the transfection. Cells maintained normal morphology during the 6-hr infection, after which they began to change shape and detach from the culture dish.

Transfection of dicistronic DNA constructs into cells

Twenty micrograms of each plasmid DNA purified by a banding in a CsCl gradient were transfected into 2 × 10^6 Huh7 cells by a calcium phosphate method (Gorman *et al.*, 1982). The cells were harvested 48 hr after transfection and cell extract was prepared by repeated freezing and thawing. The CAT assay was performed as described (Gorman *et al.*, 1982). The LacZ assay was according to Miller (1972).

RESULTS

Construction of RNAs with deletions in the 5′ UTR of the HCV genome and rationale for the method

In order to map *cis*-acting element(s) controlling translation in the HCV genome, we linked full-length (from nucleotide 1 to 341) or deleted versions of the 5′ UTR of HCV-1 RNA to the coding region of chloramphenicol acetyl transferase (CAT) mRNA (Figs. 1 and 2)

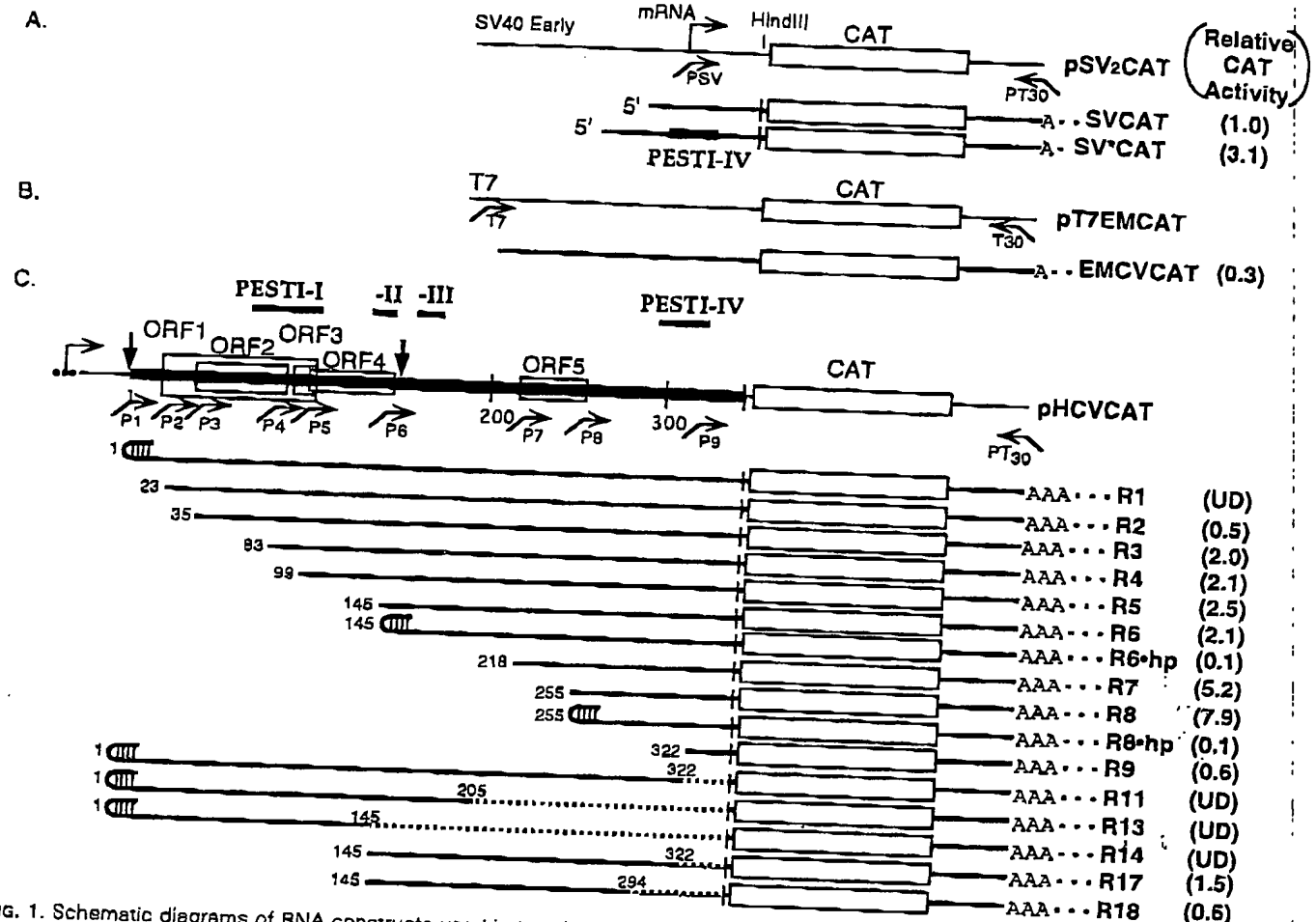


FIG. 1. Schematic diagrams of RNA constructs used in transient expression analysis. (A) SVCAT RNA. (B) EMCVCAT RNA. (C) Engineered HCVCAT RNAs (R1 to R18) with their relative CAT activities in Huh 7 cells. Position of open reading frames (ORF1 to ORF5), pestivirus homology boxes (PESTI-I to PESTI-IV), and the 5′ end of the two prominent HCV RNAs (two vertical arrows) are marked in pHCVCAT. Numbers presented in each RNA indicates the nucleotide position in the wild type HCV sequence (Han *et al.*, 1991). CAT activity is given as an average of three independent measurements. Horizontal arrows indicate primers used to amplify transcription templates. UD refers to undetectable. Dashed lines indicate 3′ deletions.

and measured CAT protein expression *in vitro* (data not shown) and CAT enzymatic activity *in vivo* (Fig. 3). We synthesized each RNA by transcribing a DNA fragment with T7 polymerase, which was first amplified by PCR to contain a specific 5′ or 3′ deletion (Fig. 1). Each RNA was designed to have a cap at the 5′ end and poly A tail (A40) at the 3′ end to increase stability in cells. This approach allows an efficient production of a large amount of RNA with uniformly defined 5′ and 3′ ends (Fig. 2). Unlike conventional DNA transfection strategies, RNA transfection of cells using this approach circumvents possible splicing and transport problems that certain RNA molecules may encounter in the nucleus. By the same method, we synthesized two additional RNAs: (1) the SVCAT with the 5′ leader of SV40 early mRNA that served as a positive control for a conventional cap-dependent translation (Kozak, 1989) and (2) the EMCVCAT with the 5′ leader of EMCV that served as a positive control for cap-independent internal initiation (Jang *et al.*, 1989).

Translation of hybrid CAT RNAs *in vitro*

In order to test whether synthetic RNAs were biologically active and to determine their translational profile *in vitro*, we translated these RNAs in rabbit reticulocyte lysate (data not shown). All RNAs including SVCAT generated a CAT protein of the expected size. The *in vitro* results can be summarized as follows: (1) In HCVCAT constructs, R1 to R5 produced CAT protein, but only at barely detectable levels. This level of translation gradually increased in R6 and in R7, reaching a

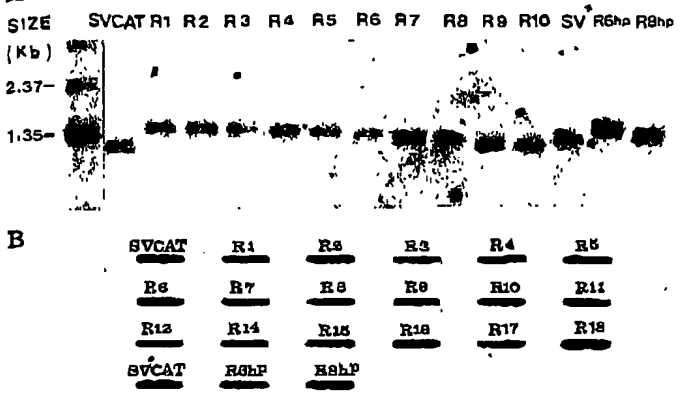


FIG. 2. Northern blot (A) and slot blot analysis (B) of engineered hybrid CAT RNAs. Two-tenths of a microgram of each denatured RNA was electrophoresed in a 1.5% agarose gel containing 0.3 *M* formamide for Northern blot analysis or directly blotted on the filter for slot blot analysis (Han *et al.*, 1986). Each blot was hybridized with a ^{32}P-labeled JHC271 which binds to the 5′ terminus of the CAT coding region.

maximum 12-fold increase in R8. (2) At K^+ concentrations of 140 m*M* the *in vitro* translation of capped SVCAT and HCVCAT RNAs (R7, R8) was more efficient than that of the uncapped RNAs by an average of 20-fold. (3) At lower K^+ concentrations (50 to 100 m*M*), translation of uncapped R1 template generated CAT protein at levels comparable to that of the capped R1 template, possibly indicating the occurrence of weak internal initiation. These results, however, did not confirm recent data by Tsukiyama-Kohara *et al.* (1992) who reported the detection of an efficient internal ribosome entry site within the 5′ UTR of HCV RNA using rabbit reticulocyte lysate and HeLa cell extracts. Because of this discrepancy and the fact that protein synthesis *in vitro* using cell lysates does not always faithfully represent translation conditions *in vivo* (Kozak, 1983), we elected to test our constructs in an *in vivo* system by transfecting mono- and dicistronic templates directly into mammalian cells as a more relevant readout of biological activity.

Translation of hybrid CAT RNAs *in vivo* and identification of control elements

In order to determine the translation profile of the monocistronic constructs *in vivo*, we transfected RNAs (R1 to R18) along with the control RNA, SVCAT, into a human hepatocyte cell line (Huh7) using lipofectin (Felgner *et al.*, 1987) and monitored CAT activities (Fig. 3). In the full-length construct R1, CAT activity was repeatedly undetectable (Fig. 3, lane 3) unless the amount of RNA was increased by 5-fold and more cell extract was used (data not shown). This result suggested that the full-length HCV RNA may not be an efficient translation template *in vivo*. When a series of 5′ deletion constructs were analyzed (Fig. 3, lanes 4–11), CAT activity was first detected in R2 in which the 5′ terminal hairpin of 23 nucleotides was removed. This activity increased by 4-fold in R3 and a similar level of activity was detected in R4, R5, and R6 which were systematically deleted for ORF1 to 4. It should be noted that the 5′ leader sequence in R6 was identical to that of the 5′ subgenomic RNA detected *in vivo* (Han *et al.*, 1991). This activity further increased by 2-fold in R7 in which the AUG codon of ORF 5 was removed and an additional 1.5-fold in R8 which retains only 86 nucleotide of 3′ proximal sequence, representing a maximum activity. These data suggested that sequences upstream from nucleotide 255 including the small ORFs are inhibitory to the translation from the major initiation codon for the polyprotein.

The maximum CAT activity seen in R8 decreased sharply upon a further deletion of 67 nucleotides (Fig. 3, R9). This result suggested that an efficient positive control element that stimulates translation may be present downstream from nucleotide 255. This 86-nucleotide region contains a 28-nucleotide sequence at position 291 to 281 with 90% sequence identity to pestiviruses and has been designated as PEST-IV (Han *et al.*, 1991). To determine whether the PEST-IV element is solely responsible for the observed translation stimulation, we performed 3′ deletion analysis on R6 (Fig. 3, lanes 12–15). We chose this RNA because any construct which contained an intact 5′ terminus of HCV RNA was inactive (see below) and 3′ deletion in R8 would generate RNA with a short 5′ leader. Upon transfection, the CAT activity seen in R6 was dropped 1.5-fold by a deletion of 20 nucleotides from the 3′ end of R6 (R17, lane 14) and a further 2.5-fold decrease by an additional deletion of 28 nucleotides (R18, lane 15). These data indicated that additional upstream and downstream sequences from the PEST-IV were neces-

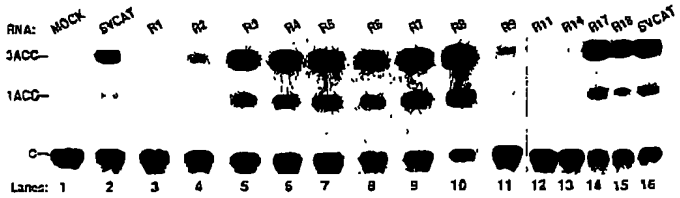


FIG. 3. Translation of transfected synthetic RNAs in Huh7 cells. Two micrograms of each RNA were transfected into 1 × 10^6 Huh7 cells using lipofectin (Felgner *et al.*, 1987) and the CAT activity was measured 14 hours later (Gorman *et al.*, 1982). The positions of unacetylated CAT (C) and monoacetylated CAT products (1ACC or 3ACC) are indicated.

sary for maximum translational enhancement. Nonetheless, this PEST-IV sequence appears to be a part of a positive *cis*-acting element which can be transferred to a heterologous 5′ leader: when this 28-nucleotide sequence was inserted into the SVCAT to create SV*CAT, it conferred an increase in CAT activity of 3-fold (Fig. 4, lane 2 vs lane 3).

In additional 3′ deletion analysis, no CAT activity was detected in R11 and R14 (Fig. 3, lanes 12, 13) or in R13 (data not shown), all of which contained the 5′ hairpin. These data are consistent with the view that the 5′ hairpin may be inhibitory to the translation of HCV RNA.

The effect of 5′ hairpin of HCV on the translation of CAT RNAs

Since RNAs with the intact 5′ terminus were all inactive irrespective of the downstream sequences (Fig. 3, lanes 3, 12, 13), we tested whether a potential 5′ hairpin structure resident in the most distal 23 nucleotides (hereafter referred to as the 5′ hairpin) is directly responsible for the observed translation inhibition. Accordingly, the 5′ hairpin was linked to the 5′ terminus of the two active RNAs, R6 and R8. These RNAs (R6hp, R8hp) were transfected into Huh7 cells and the CAT activity was measured. As shown in Fig. 4, the juxtaposition of the hairpin on these constructs nearly abolished the translation as demonstrated by the relative CAT activity (lane 7 vs lane 9, lane 10 vs lane 12). This indicates that the 5′ hairpin is a potent translation inhibitor, although complete inhibition requires more sequence than the 23-nucleotide hairpin alone.

Translation of hybrid CAT RNAs in poliovirus-infected cells

Poliovirus infection is known to inhibit the cap-dependent translation of cellular mRNA and thereby promote translation of its own or heterologous RNA which contains an IRES within its 5′ UTR (Jang *et al.*, 1989; Macejak and Sarnow, 1991; Pelletier and Sonenberg, 1988). This inhibition is believed to be mediated indirectly by the poliovirus-encoded proteinase 2A by activating an unidentified latent cellular protease which in turn cleaves p220, a component of the cellular cap-binding protein complex (eIF–4F) (Sonenberg, 1988). Therefore, we transfected hybrid CAT RNAs with various 5′ UTR into Huh7 cells infected with poliovirus. This strategy was designed to determine the cap dependency of each RNA and to detect the possible existence of a weak IRES which may be present in the HCV 5′ UTR of HCV-1 RNA. As expected, poliovirus infection increased CAT activity by sevenfold in EMCVCAT (Fig. 4, lane 13 vs lane 14), a positive control RNA for internal initiation (Elroy-Stein *et al.*, 1989). In contrast, the

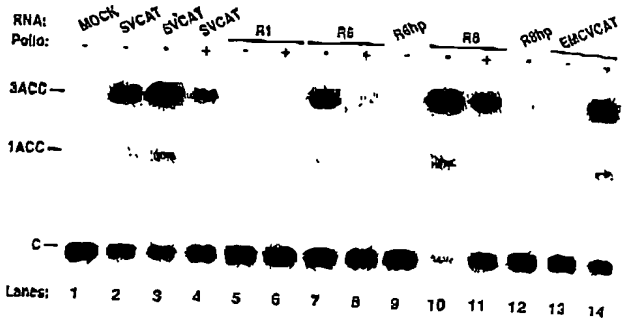


FIG. 4. The effect of poliovirus infection, the 5′ hairpin, and the PEST-IV element on CAT activity in Huh7 cells. Cells (1×10^6) were uninfected or infected (lane 4, 8, 11, and 14) with poliovirus type 1 (ATCC VR-59) 2 hr before RNA transfection at a m.o.i. of 100. Transfection conditions were the same as in Fig. 3 except that CAT activity was assayed 4 hr after RNA transfection.

infection substantially decreased CAT activity in SVCAT (lane 2 vs lane 4) as well as in two HCV constructs, R6 (lane 7 vs lane 8) and R8 (lane 10 vs lane 11), respectively. The lowered CAT activities seen in poliovirus-infected cells were further diminished if infected cells were incubated longer than 2.5 hr prior to RNA transfection (data not shown). The CAT activity of R1 remained undetectable regardless of poliovirus infection (lanes 5, 6). This result strongly suggested that an IRES is not present in the 5′ UTR of HCV-1 RNA.

With the exception of the R1 construct, the constructs tested in the above experiment contained large deletions of the 5′ UTR. Because it is formally possible that such deletions may have affected a putative IRES structure and/or function, we elected to test constructs with less extensive deletions for their ability to translate CAT protein in poliovirus-infected cells. In agreement with results shown in Fig. 4, the CAT activity of the R1 construct remained undetectable in the presence or absence of poliovirus infection (Fig. 5A, lanes 3 and 4). Constructs R2 to R5 showed relative levels of CAT activity similar to those described previously (e.g., Fig. 3) when tested in the absence of poliovirus infection (Fig. 4A, lanes 5, 7, 9, and 11, respectively). However, the CAT activities of the HCV leader templates were practically abolished in the poliovirus-infected cells (Fig. 5A, lanes 6, 8, 10, and 12).

In addition, templates SVCAT and R1 to R3 were tested in a similar protocol using uncapped messages. As is shown in Fig. 5B, the uncapped templates were inactive in transfected cells whether or not the cells were subsequently infected with poliovirus. These results strongly suggest that monocistronic messages with *cis*-acting regulatory elements derived from the HCV 5′ UTR are translated by a cap-dependent mecha-

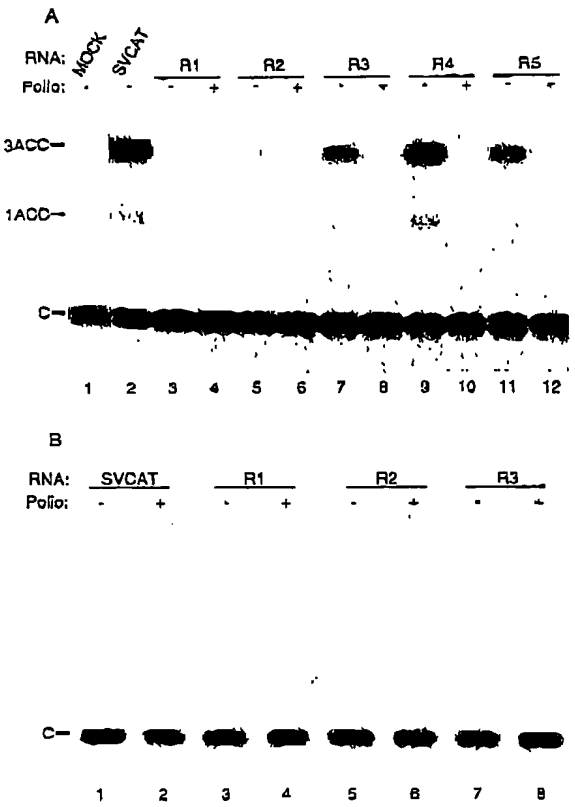


FIG. 5. The effect of poliovirus infection on CAT activity of capped (A) or uncapped (B) RNAs in Huh7 cells. Cells (1×10^6) were uninfected or infected (lanes 4, 6, 8, 10, and 12 for A; lanes 2, 4, 6, and 8 for B) with poliovirus, transfected with indicated RNAs, and assayed for CAT activity. Experimental conditions were the same as in Fig. 4.

nism and that the HCV 5'-noncoding region does not have an IRES element.

Translation of dicistronic mRNA in Huh7 cells

The possible presence of an IRES within the 5' leader of HCV was further tested by transfecting Huh7 cells with DNA constructs designed to transcribe a dicistronic mRNA. Thus we placed the 5' UTR of HCV RNA as an intercistronic spacer between CAT as the first cistron and LacZ as the second cistron and cloned this linked DNA into an expression vector, in which transcription is derived by the strong enhancer-promoter of the major immediate–early gene in cytomegalovirus (CMV) (Fig. 6A). In addition, we constructed both positive and negative control dicistronic vectors, in which the 5' UTR of HCV was replaced with the 5' UTR of poliovirus and the 3' UTR of SV40 early gene, respectively. Upon transfection into Huh7 cells, all three constructs supported translation of the first CAT cistron at a comparable level (Fig. 6B); however, the dicistronic construct with the HCV leader did not support the translation of the second LacZ cistron at a level comparable to the dicistronic control construct employing a poliovirus leader (Fig. 6C). These data support the earlier evidence generated using monocistronic constructs that the full-length 5' UTR of HCV genome does not contain an IRES.

Translation of HCV RNA constructs in HeLa and HepG2 cells

Transient transfection assays can give different readouts that are cell line dependent. In order to ensure that the results we obtained were not confined to HUH7 cells, we transfected R1, R7, and R8 constructs into HeLa and HepG2 cells to assay the constructs in different cell lines. The resultant pattern of CAT activity were qualitatively similar to that observed in Huh7 cells (Fig. 7).

DISCUSSION

HCV is believed to be a distant relative of flaviviruses (Choo *et al.*, 1989, 1990; Han *et al.*, 1991; Houghton *et al.*, 1991), but its genome has structural features at the 5' and 3' termini shared with that of poliovirus (Kitamura *et al.*, 1981) in two respects. First, the 5' UTR of HCV is relatively long, contains multiple ORFs, can be modeled into a highly ordered structure (data not shown), and has a putative hairpin structure at the 5' terminus (Han *et al.*, 1991; Inchauspe *et al.*, 1991; Okamoto *et al.*, 1991, 1992). Second, the 3' UTR of HCV is short and has a homopolymer tail (Han and Houghton, 1992). Since the 5' UTR of poliovirus RNA is translated by a cap-independent internal initiation mechanism (Pelletier and Sonenberg, 1988), we searched for the same mechanism in the 5' UTR of HCV-1 RNA. However, we failed to detect such an activity *in vivo* in conjunction with poliovirus infection. In addition, we could not detect evidence for internal initiation using a dicistronic mRNA approach, a standard DNA transfection system for the detection of IRES in a test RNA (Jang *et al.*, 1989; Pelletier and Sonenberg, 1988). Our results are not cell type-specific, since we obtained similar results from several human cell lines including HeLa and HepG2 cells (data not shown). Taken together, we conclude that the 5' UTR of HCV-1 RNA does not contain an IRES. Furthermore, the fact that all test RNAs having the hairpin at the 5' terminus were translationally inactive strongly suggests that translation of HCV RNA is 5' end-dependent.

In contrast to our results from HCV-1, a group I isolate, Tsukiyama-Kohara *et al.* (1992) recently reported the detection of IRES within the 5' UTR of HCV RNA from two Japanese isolates which belong to group II

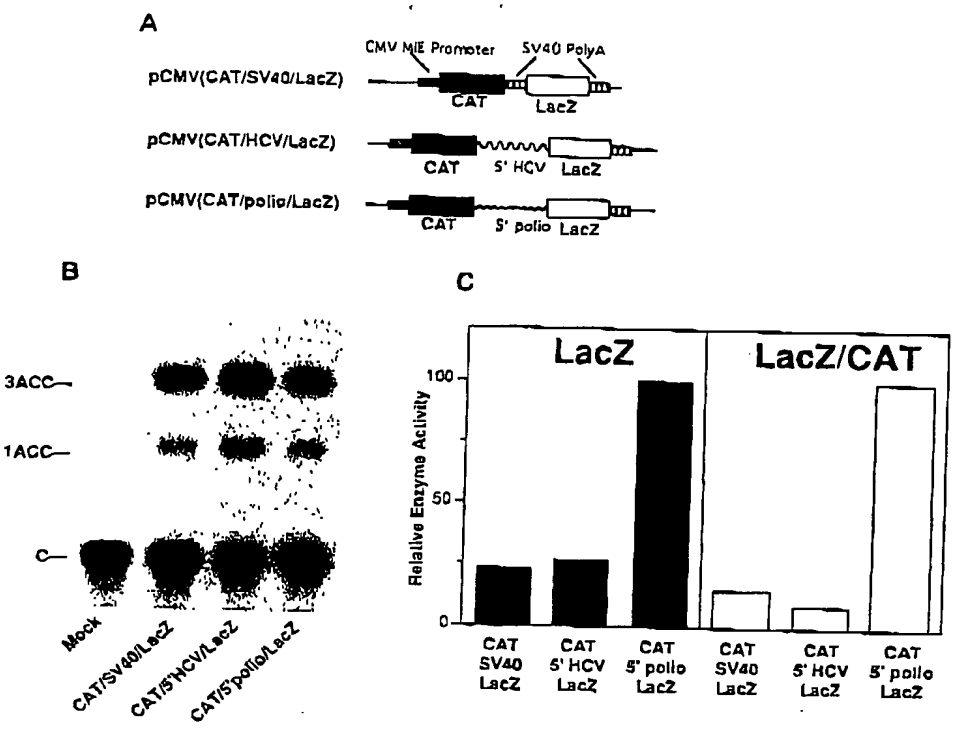


FIG. 6. Expression of CAT and LacZ from dicistronic mRNAs. (A) Schematic representation of expression vectors for each mRNA. In each plasmid, the CMV major immediate-early promoter (MIE) and SV40 polyadenylation signal are indicated. (B) CAT activity from each dicistronic mRNA. (C) Relative LacZ activity normalized to CAT activity.

and III, respectively. Currently, it is difficult to explain these discrepant results. Although it is less likely, it may be formally possible that different groups of HCV have adapted different translation strategies during the course of evolution. The two Japanese HCV isolates vary in the strength of internal initiation by twofold (Tsukiyama-Kohara *et al.*, 1992). The nucleotide sequence in the 5′ UTR of the two reported Japanese HCV isolates differs from each other by 5% and from HCV-1 by 3 and 6%, respectively. Among HCV isolates including these three, the sequence heterogeneity within the 5′ UTR is mainly clustered between nucleotide 200 and 250, which may form a potential secondary structure. In this structure, a pyrimidine track which may be important for internal initiation (Luz and Beck, 1991; Pestova *et al.*, 1991) is located within a loop region (Tsukiyama-Kohara *et al.*, 1992). Although HCV-1 has this sequence in the same region, our results do not indicate that the presence of it confers a cap-independent translation phenotype. Alternatively, *in vitro* systems can be influenced in their cap dependence by levels of K^+ ion in the lysate (Herman, 1987). It is not known what K^+ ion levels were present in the lysates used in the earlier report by Tsukiyama-Kohara *et al.* (1992).

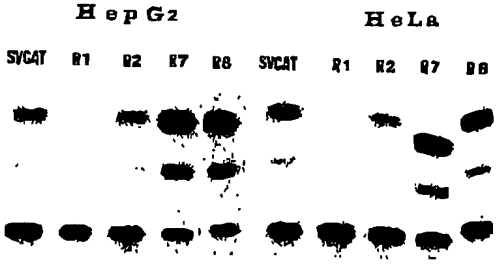


FIG. 7. Translation of hybrid CAT RNAs transfected into HepG2 cells and HeLa cells. RNA transfection and CAT assay were performed as described in Fig. 3 legend.

At the 5′ terminus, HCV-1 RNA has a 27-nucleotide sequence that has a potential to form a hairpin structure with calculated free energy of −14.5 kcal/mol (Han *et al.*, 1991). Although the existence of this hairpin in physiological conditions needs to be verified by nuclease digestion experiments, conservation of its structure, but not the primary sequence, in all putative full-length HCV sequences reported to date (Chen *et al.*, 1992; Han *et al.*, 1991; Inchauspe *et al.*, 1991; Okamoto *et al.*, 1991, 1992) implies that the putative hairpin could be functionally important. We proved that sequence residing within this proposed 5′ hairpin functions as a potent translational inhibitor. It was shown that a stable secondary structure in the 5′ UTR of

mRNA ($\Delta G = -50$ kcal/mol) reduces translation efficiency (Kozak, 1986). By comparison, the putative 5′ hairpin of HCV has a relatively weak secondary structure, which may not be sufficient to block the unwinding activity of the initiation complex (Sonenberg, 1988). However, because of its location, it may block ribosome entry onto the 5′ end of the RNA. Alternatively, it may be a binding site for cellular factor(s) which could be involved in viral replication or encapsidation. RNA hairpins have been implicated in translational control of cellular and viral mRNAs. In ferritin mRNA, a hairpin in the 5′ UTR reversibly binds a cytosolic protein in the absence of iron, which results in translation repression (Rouault *et al.*, 1988). In *Xenopus* oocytes, introduction of an artificial hairpin into the 5′ terminus of a test mRNA results in regulation of translation, which is specific to the stage of differentiation (Fu *et al.*, 1991). A 5′ proximal hairpin is highly conserved between enteroviruses and rhinoviruses. In poliovirus, the RNA hairpin is required for efficient translation and viral replication and mutation in this region significantly lowers translation efficiency, suggesting that the hairpin potentiates the internal ribosome initiation process possibly by its interaction with downstream element or protein factor(s) (Simoes and Sarnow, 1991). Thus the role in translation played by the putative 5′ hairpin of the HCV genome is opposite to that of the poliovirus genome, a marked difference between HCV and poliovirus.

We have located a putative *cis*-acting element that efficiently enhances translation within an 86-nucleotide region, which contains the PEST-IV homology box. The latter 28-nucleotide sequence is nearly perfectly conserved among HCV isolates and shares 90% nucleotide sequence identity with pestiviruses (Han *et al.*, 1991), implying a functional significance. Although we have not determined the precise 5′ and 3′ border of the positive element which enhances translation, we have demonstrated that the PEST-IV homology box is a part of such an element. Currently, it is unknown how this element augments translation *in vivo*. Based on its relatively short sequence requirement, it may facilitate translation by providing a higher relative affinity for limiting component(s) of translation machinery as suggested for alfalfa mosaic virus (Jobling and Gehrke, 1987).

The upstream ORFs, especially the last three of the five, are conserved in all HCV isolates suggesting their possible role in translation. However, defining a role for the first four ORFs in translational control is complicated by two structural features: (1) the first ORF is a part of the 5′ hairpin and (2) three of the remaining four ORFs are overlapping. A deletion in ORF1 (R2) resulted in a moderate increase in CAT activity which may be explained by disruption of the inhibitory element at the 5′ terminus. Removal of 35 nucleotides including the second AUG (R3) resulted in a fourfold increase in CAT activity. Subsequent deletion of the remaining ORFs (R4–R6) did not change CAT activity. Deletion of AUG codon in the fifth ORF (R7) results in a 2.5-fold increase in CAT activity. These data suggest that the upstream ORFs may function as negative modulators of translation. This is consistent with the ribosome scanning hypothesis that has been reported to be operative for the majority of cellular and viral mRNAs (Kozak, 1983, 1989).

We have mapped three putative control elements in the 5′ UTR of HCV-1 RNA. However, the fact that the full-length HCV-1 RNA is translationally inactive raises a question as to how HCV initiates polyprotein synthesis upon infection. In the absence of experimental evidence from a cell culture system, one can consider several possibilities. Perhaps HCV infection primes initial translation by viral component(s). This could either be an HCV-encoded factor(s) which derepresses the translational inhibition imposed by the 5′ repressor element in the genomic RNA or an as yet unidentified *cis*-acting element(s) elsewhere in its genome. downstream from polyprotein initiation codon, which allows internal initiation. However, in view of the arrangement of the repressor element at the 5′ end and two putative *cis*-acting elements within the 5′ UTR of HCV, we propose that active mRNA is present in infected cells as a separate entity. Previously, we reported the detection of both 5′ and 3′ subgenomic RNAs (Han *et al.*, 1991). Although the biological significance of these RNAs has not been established, we believe that the 5′ subgenomic RNA could possibly be a viral mRNA based on the fact that a CAT RNA having the 5′ leader of this RNA is translationally active. Currently, the origin of this 5′ subgenomic RNA is unknown. One possibility is that it could be transcribed from a specific promoter element within the 5′ leader by RNA dependent RNA polymerase, for example, as described in Sindbis virus (Strauss and Strauss, 1986).

We speculate that viral protein synthesis in infected cells may be regulated at two or more levels, including mRNA production and control at the level of translation. HCV is believed to exist at low titer in clinical samples. However, HCV infection is persistent and leads to chronic hepatitis and hepatocellular carcinoma at unusually high frequencies (Dienstag and Alter, 1986; Houghton *et al.*, 1991). We hypothesize that viral replication and translation control operating at various levels within the 5′ leader of HCV may be a genetic mechanism, in conjunction with a possible immunologic mechanism (Weiner *et al.*, 1992) for the observed pathobiology of viral infection. Our findings of transla-

tional control elements in HCV RNA provide intriguing prospects for future HCV research.

ACKNOWLEDGMENTS

We thank Dr. V. Shyamala for the preparation of recombinant CAT RNAs; Dr. P. Barr for the synthesis of oligonucleotides; C. Degnin for technical assistance; E. Lim, D. Alexander, and G. Sequar for construction and preparation of plasmid DNAs; S. Klein for sequencing; Dr. B. Yen for the supply of Huh7 cells; Dr. C. S. Hahn for the helpful discussion on RNA transfection; Dr. T. Sanders for a critical review; and T. Calarco for graphics. This work was supported by Chiron, CIBA-Geigy, and Ortho Diagnostics system and by NIH Grant AI26672 awarded to A. P. Geballe.

REFERENCES

ADAM, M. A., RAMESH, N., MILLER, A. D., and OSBORNE, W. R. (1991). Internal initiation of translation in retroviral vectors carrying picornavirus 5' nontranslated regions. *J. Virol.* 65, 4985–4990.

ALTER, H. J., PURCELL, R. H., SHIH, J. W., MELPOLDER, J. C., HOUGHTON, M., CHOO, Q.-L., and KUO, G. (1989). Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. *N. Engl. J. Med.* 321, 1494–1500.

CHEN, E. Y., and SEEBURG, P. H. (1985). Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. *DNA* 4, 165–170.

CHEN, P.-J., LIN, M., TAI, K.-F., LIU, P.-C., LIN, C.-J., and CHEN, D.-S. (1992). The Taiwanese hepatitis C virus genome: Sequence determination and mapping the 5' terminus of viral genoma and antigenomic RNA. *Virology* 88, 102–113.

CHOO, Q.-L., KUO, G., WEINER, A. J., OVERBY, L. R., BRADLEY, D. W., and HOUGHTON, M. (1989). Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* 244, 359–362.

CHOO, O.-L., WEINER, A. J., OVERBY, L. R., KUO, G., HOUGHTON, M., and BRADLEY, D. W. (1990). Hepatitis C virus: The major causative agent of viral non-A, non-B hepatitis. *Br. Med. Bull.* 46, 423–441.

CHOO, Q.-L., RICHMAN, K. H., HAN, J. H., BERGER, K., LEE, C., DONG, C., GALLEGOS, C., COIT, D., MEDINA-SELBY, A., BARR, P. J., WEINER, A. J., BRADLEY, D. W., KUO, G., and HOUGHTON, H. (1991). Genetic organization and diversity of the hepatitis C virus. *Proc. Natl. Acad. Sci. USA* 88, 2451–2455.

DIENSTAG, J. L., and ALTER, H. J. (1986). Non-A, non-B hepatitis: Evolving epidemiologic and clinical perspective. *In* "Seminars in Liver Disease," pp. 67–91. Thieme, New York.

ELROY-STEIN, O., FUERST, T. R., and MOSS, B. (1989). Cap-dependent translation of mRNA conferred by encephalomyocarditis virus 5' sequence improves the performance of vaccinia virus/bacteriophage T7 hybrid expression system. *Proc. Natl. Acad. Sci. USA* 86, 6126–6130.

FELGNER, P. L., GADEK, T. R., HOLM, M., ROMAN, R., CHAN, H. W., WENZ, M., NORTHROP, J. P., RINGOLD, G. M., and DANIELSON, M. (1987). Lipofection: A highly efficient, lipid-mediated DNA-transfection procedure. *Proc. Natl. Acad. Sci. USA* 84, 7413–7417.

FU, L., YE, R., BROWDENER, L. W., and JOHNSTON, R. N. (1991). Translational potentiation of mRNA with secondary structure in Xenopus. *Science* 251, 807–810.

GORMAN, C. M., MOFFAT, L. F., and HOWARD, B. H. (1982). Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* 2, 1044–1051.

HAN, J. H., SHYAMALA, V., RICHMAN, K. H., BRAUER, M. B., IRVINE, B., URDEA, M. S., TEKAMP-OLSON, P., KUO, G., CHOO, Q.-L., and HOUGHTON, M. (1991). Characterization of the terminal regions of hepatitis C viral RNA: Identification of conserved sequences in the 5' untranslated region and poly(A) tails at the 3' end. *Proc. Natl. Acad. Sci. USA* 88, 1711–1715.

HAN, J. H., RALL, L., and RUTTER, W. J. (1986). Selective expression of rat pancreatic genes during embryonic development. *Proc. Natl. Acad. Sci. USA* 83, 110–114.

HAN, J. H. and HOUGHTON, M. (1992). Group specific sequence and conserved secondary structures at the 3' end of HCV genome and its implication for viral replication. *Nucl. Acids Res.* 20, 3520.

HERMAN, R. C. (1987). Characterization of the internal initiation of translation on the vesicular stomatitis virus phosphoprotein mRNA. *Biochemistry* 26, 8346–8350.

HOUGHTON, M., WEINER, A., HAN, J., KUO, G., and CHOO, Q.-L. (1991). Molecular biology of the hepatitis C viruses: Implication for diagnosis, development and control of viral disease. *Hepatology* 14, 381–388.

INCHAUSPE, G., ZEVEDEE, S., LEE, D.-H., SUGITANI, M., NASOFF, M., and PRINCE, A. M. (1991). Genome structure of the human prototype strain H of hepatitis C virus: Comparison with American and Japanese isolates. *Proc. Natl. Acad. Sci. USA* 88, 10292–10296.

JANG, S. K., DAVIS, M., KAUFMAN, J., and WIMMER, E. (1989). Initiation of protein synthesis by internal entry of ribosomes into the 5' nontranslated region of Encephalomyocarditis virus RNA *In vivo*. *J. Virol.* 63, 1651–1660.

JOBLING, S. A., and GEHRKE, L. (1987). Enhanced translation of chimeric mRNAs containing a plant viral untranslated leader sequence. *Nature* 325, 623–625.

KATO, N., HUIKATA, M., OOTSUYAMA, Y., NAKAGAWA, M., OHKOSHI, S., SUGIMURA, T., and SHIMOTOHNO, K. (1990). Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proc. Natl. Acad. Sci. USA* 87, 9524–9529.

KITAMURA, N., SEMLER, B., ROTHBERG, P. G., LARSEN, G. R., ADLER, C. J., DORNER, A. J., EMINE, E. A., HANECAK, R., LEE, J. L., WERF, S., ANDERSON, C. W., and WIMMER, E. (1981). Primary structure, gene organization and polypeptide expression of poliovirus RNA. *Nature* 291, 547–553.

KOZAK, M. (1983). Comparison of initiation of protein synthesis in prokaryotes, eukryotes, and organells. *Microbiol. Rev.* 47, 1–45.

KOZAK, M. (1986). Influence of a mRNA secondary structure on initiation by eukaryotic ribosome. *Proc. Natl. Acad. Sci. USA* 83, 2850–2854.

KOZAK, M. (1989). The scanning model for translation: An update. *J. Cell Biol.* 108, 229–241.

KUO, G., CHOO, Q.-L., ALTER, H. J., GITNICK, G. L., REDEKER, A. G., PURCELL, R. H., MIYAMURA, T., DIENSTAG, J. L., ALTER, M. J., STEVENS, C. E., TEGTMEIER, G. E., BONINO, F., COLOMBO, M., LEE, W.-S., KUO, C., BERGER, K., SHUSTER, J. R., OVERBY, L. R., BRADLEY, D. W., and HOUGHTON, M. (1989). An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science* 244, 362–364.

LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685.

LUZ, N., and BECK, E. (1991). Interaction of a cellular 57-kilodalton protein with the internal translation initiation site of foot-and-mouth disease virus. *J. Virol.* 65, 6486–6494.

MACEJAK, D. G., and SARNOW, P. (1991). Internal initiation of translation mediated by the 5' leader of a cellular mRNA. *Nature* 353, 90–94.

MILLER, J. H. (1972). Assay of beta-galactosidase. *In* "Experiments in Molecular Genetics," pp. 352–355. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

MORI, S., KATO, N., YAGYU, A., TANAKA, T., IKEDA, Y., PETCHCLAI, B., CHIEWSILP, P., KURIMURA, T., and SHIMOTOHNO, K. (1992). A new

type of hepatitis C virus in Thailand. *Biochem. Biophys. Res. Commun.* 183, 334–342.

OKAMOTO, H., OKADA, S., SUGIYAMA, Y., KURAI, K., IIZUKA, H., MACHIDA, A., MIYAKAWA, Y., and MAYUMI, M. (1991). Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: Comparison with reported isolates for conserved and divergent regions. *J. Gen. Virol.* 72, 2697–2704.

OKAMOTO, H., KURAI, K., OKADA, S., YAMAMOTO, K., LIZUKA, H., TANAKA, T., FUKUDA, S., TSUDA, F., and MISHIRO, S. (1992). Full-length sequence of hepatitis C virus genome having poor homology to reported isolates: Comparative study of four distinct genotypes. *Virology* 188, 331–341.

PELLETIER, J., and SONENBERG, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 334, 320–325.

PESTOVA, T. V., HELLEN, C. U. T., and WIMMER, E. (1991). Translation of poliovirus RNA: Role of an essential cis-acting oligopyrimidine element within the 5′ nontranslated region and involvement of a cellular 57-kilodalton protein. *J. Virol.* 65, 6194–6204.

ROUAULT, T. A., HENTZE, M. W., CAUGHMAN, S. W., HARFORD, J. B., and KLAUSNER, R. D. (1988). Binding of a cytosol protein to the iron-responsive element of human ferritin mRNA. *Science* 241, 1207–1210.

SAIKI, R. K., GELFAND, D. H., STOFFEL, S., SCHARF, S. J., HIGUCHI, R., HORN, G. T., MULLIS, K. B., and ERLICH, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487–491.

SCHLEISS, M. R., DEGNIN, C. R., and GEBALLE, A. P. (1991). Translational control of human cytomegalovirus gp48 expression. *J. Virol.* 65, 6782–6789.

SHYAMALA, V., and AMES, G. F.-L. (1991). Use of exonuclease for rapid polymerase chain reaction based *in vitro* mutagenesis. *Gene* 97, 1–6.

SIMOES, E. A., and SARNOW, P. (1991). An RNA hairpin at the extream 5′ end of the poliovirus RNA genome modulates viral translation in human cells. *J. Virol.* 65, 913–921.

SONENBERG, N. (1988). Cap-binding protein of eukaryotic mRNA: Functions in initiation and control of translation. *Prog. Nucleic Acid Res. Mol. Biol.* 35, 173–297.

STRAUSS, E., and STRAUSS, J. H. (1986). Structures and replication of the alphavirus genome. *In* "Togaviridae and Flaviviridae" (S. Schlesinger and M. J. Schlesinger, Eds.), pp. 347–349. Plenum. New York.

TAKAMIZAWA, A., MORI, C., FUKE, I., MANABE, S., MURAKAMI, S., FUJITA, J., ONISHI, E., ANDOH, T., YOSHIDA, I., and OKAYAMA, H. (1991). Structure and organization of the hepatitis C virus genome isolated from human carriers. *J. Virol.* 65, 1105–1113.

TANAKA, T., KATO, N., NAKAGAWA, M., OOTSUYAMA, Y., CHO, M. J., NAKAZAWA, T., HIJIKATA, M., ISHIMURA, Y., and SHIMOTOHNO, K. (1992). Molecular cloning of hepatitis C virus genome from a single Japanese carrier: Sequence variation within the same individual and among infected individuals. *Virus Res.* 23, 39–53.

TSUKIYAMA-KOHARA, K., IIAUKA, N., KOHARA, M., and NOMOTO, A. (1992). Internal ribosome entry site within hepatitis C virus RNA. *J. Virol.* 66, 1476–1483.

WEINER, A. J., GEYSEN, H. M., CHRISTOPHERSON, C., HALL, J. E., MASON, T. J., SARACCO, G., BONINO, F., CRAWFORD, K., MARION, C. D., CRAWFORD, K. A., BARR, P. J., BRUNETTO, M., MIYAMURA, T., MCHUTCHINSON, J., and HOUGHTON, M. (1992). Theoretical and empirical evidence for immune selection of HCV E2/NS1 glycoprotein variants: putative role in chronic hepatitis C virus infections. *Proc. Natl. Acad. Sci. USA* 89, 3468–3472.

Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 4942–4946, June 1992
Biochemistry

Sequence analysis of the 5′ noncoding region of hepatitis C virus

(non-A, non-B hepatitis/genetic heterogeneity/virus evolution)

JENS BUKH, ROBERT H. PURCELL*, AND ROGER H. MILLER

Hepatitis Viruses Section, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892

Contributed by Robert H. Purcell, February 14, 1992

ABSTRACT **We have determined the nucleotide sequence of the 5′ noncoding (NC) region of the hepatitis C virus (HCV) genome in 44 isolates from around the world. We have identified several HCV isolates with significantly greater sequence heterogeneity than reported previously within the 5′ NC region. The most distantly related isolates were only 90.1% identical. Nucleotide insertions were seen in three isolates. Analysis of the nucleotide sequence from 44 HCV isolates in this study combined with that of 37 isolates reported in the literature reveals that the 5′ NC region of HCV consists of highly conserved domains interspersed with variable domains. The consensus sequence was identical to the prototype HCV sequence. Nucleotide variations were found in 45 (16%) of the 282 nucleotide positions analyzed and were primarily located in three domains of significant heterogeneity (positions −239 to −222, −167 to −118, and −100 to −72). Conversely, there were three highly conserved domains consisting of 18, 22, and 63 completely invariant nucleotides (positions −263 to −246, −199 to −178, and −65 to −3, respectively). Two nucleotide domains within the 5′ NC region, conserved among all HCV isolates studied to date, shared statistically significant similarity with pestivirus 5′ NC sequences, providing further evidence for a close evolutionary relationship between these two groups of viruses. Additional analysis revealed the presence of short open reading frames in all HCV isolates. Our sequence analysis of the 5′ NC region of the HCV genome provides additional information about conserved elements within this region and suggests a possible functional role for the region in viral replication or gene expression. These data also have implications for selection of optimal primer sequences for the detection of HCV RNA by the PCR assay.**

The etiological agent of most posttransfusion non-A, non-B hepatitis cases, hepatitis C virus (HCV), is a positive-stranded RNA virus with a linear genome ≈9.5 kilobases (kb) in length (1, 2). The structure and organization of the HCV genome are similar to that of pesti- and flaviviruses (3–5). Published sequence data indicate that the 5′ noncoding region (NC) of 324–341 nucleotides is generally highly conserved among different HCV isolates (6–8) and, furthermore, is the most highly conserved region of the HCV genome (3–5, 9, 10). Also the 5′ NC region of the HCV genome shares sequence similarity with the 5′ NC region of pestiviruses (5, 7). The high degree of sequence conservation has made this region the choice for primer selection in reverse transcription and amplification of HCV RNA by PCR (cDNA PCR) (for review, see ref. 11). Four short open reading frames (ORFs) have been described in the 5′ NC region of HCV (5, 7), but the significance of these ORFs is unknown. Overall, these findings suggest an important functional role of the 5′ NC region of the HCV genome in virus replication or gene expression.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

In a recent study we tested sera from 114 individuals positive for antibodies to HCV (anti-HCV) from around the world for the presence of HCV RNA with four different primer sets in a cDNA PCR assay (12). In this study we have determined the nucleotide sequence of the 5′ NC region of the HCV genome of 44 of these HCV isolates.† The isolates selected for analysis were chosen because they are representative of the different geographical locations and of the different patterns of reactivity to the primers used in the previous study (12). We find that, contrary to previously published data, the 5′ NC region of the HCV genome possesses significant heterogeneity among different HCV isolates.

MATERIALS AND METHODS

Serum samples used in this study were from 44 anti-HCV-positive individuals from 12 countries [Denmark (DK), Dominican Republic (DR), Germany (D), Hong Kong (HK), India (IND), Italy (S), Peru (P), South Africa (SA), Sweden (SW), Taiwan (T), United States (US), and Zaire (Z)]. These samples were used in a recent study in which sera from 114 anti-HCV-positive individuals were tested for HCV RNA in a cDNA PCR assay with four primer sets (12). Primer set *a* was from within the 5′ NC region of the HCV genome; primer set *b* spanned the 5′ end of the 5′ NC region to the 5′ end of the putative core gene sequence; primer set *c* was from the 3′ end of the 5′ NC region to the 3′ end of the core gene region; and primer set *d* was from the nonstructural protein 3-like gene of HCV. We selected for sequence analysis HCV isolates that represented each of the 12 countries and that reflected probable heterogeneity as measured by the different patterns of reactivity with primer sets *a–d*. Specifically, isolates DR4, DK7, HK5, S9, SW2, T3, and US11 were positive with primer sets *a*, *b*, *c*, and *d*; isolates D3, D6, DK11, DK13, IND8, P10, SA1, SA7, SA10, SW3, T10, US6, and Z4 were positive with primer sets *a*, *b*, and *c*; isolate DK9 was positive with primer sets *a*, *b*, and *d*; isolates DK12, HK2, HK10, IND3, IND5, P8, S45, S52, S54, S83, SA11, T8, T9, US3, Z1, Z5, Z6, and Z8 were positive with primer sets *a* and *b*; and isolates SA3, T4, US1, US10, and Z7 were positive only with primer set *a*. Viral RNA was extracted from serum, reverse-transcribed, and the resulting cDNA was amplified in a nested PCR assay as described (12). For 39 HCV isolates that were detected with primer set *b* we sequenced that PCR product, a 321-nucleotide DNA fragment that spanned 282 nucleotides of the 5′ NC region and 39 nucleotides of the core gene region of HCV (i.e., positions −282 to 39). In five HCV isolates that could be amplified only with primer set *a*, that PCR product, a 196-nucleotide DNA fragment from the 5′ NC region (i.e., positions −246 to −51)

Abbreviations: HCV, hepatitis C virus; NC, noncoding; ORF, open reading frame; BVDV, bovine viral diarrhea virus.

*To whom reprint requests should be addressed.

†The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M84822–M84865).

was sequenced. A number of standard precautions were taken to reduce the risk of contamination with exogenous RNA, and a negative control was included for every serum sample tested in the RNA extraction, reverse transcription, and PCR amplification to rule out contamination as a source of false positive results (12). Amplified DNA for sequencing was purified by gel electrophoresis as described (12), and ≈100 ng of DNA was used for direct sequencing by the dideoxynucleotide chain-termination method (13, 14) with phage T7 DNA polymerase (Sequenase, United States Biochemical). Serum containing the prototype HCV isolate (HCV-1, refs. 5 and 7), provided by D. W. Bradley (Centers for Disease Control, Atlanta), was used as a positive control in PCR and sequencing reactions. Computer analysis of the sequences of HCV, flavivirus, and pestivirus genomes was done as described (15, 16).

RESULTS AND DISCUSSION

Nucleotide Sequence of the 5′ Noncoding Region from 44 HCV Isolates. A primary goal of this investigation was to analyze the nucleotide sequence of the 5′ NC region of the HCV genome from a large number of isolates obtained throughout the world. Therefore, we reverse-transcribed HCV RNA, PCR-amplified the resultant cDNA, and directly sequenced the product to obtain the "consensus" sequence in each serum sample. An alignment of the nucleotide sequence of the 5′ NC region from the 44 HCV isolates we studied is presented in Fig. 1. Previously published multiple sequence alignments of others (6–8) demonstrated that the 5′ NC region was highly conserved and 98% identical with prototype isolate HCV-1. In contrast, we have identified several HCV isolates with significantly more sequence variability. (*i*) We found that three isolates had nucleotide insertions. Isolate HK2 had two separate nucleotide insertions of one and two nucleotides, whereas isolates Z5 and Z8 each had a single nucleotide insertion. We did not detect nucleotide deletions. (*ii*) Nucleotide variation among all of the additional HCV isolates was as high as 9.9% (HK10, S52, S54 versus DK11, T8, SW3), and the nucleotide variation from HCV-1 was as high as 6.4% (DK12, HK10, S52, and S54) within the region of 282 nucleotides sequenced. Thus, contrary to the findings of others, we have demonstrated significant sequence variation within the 5′ NC region of the HCV genome.

Houghton and coworkers (11) analyzed the degree of sequence heterogeneity of HCV isolates and, based on this,

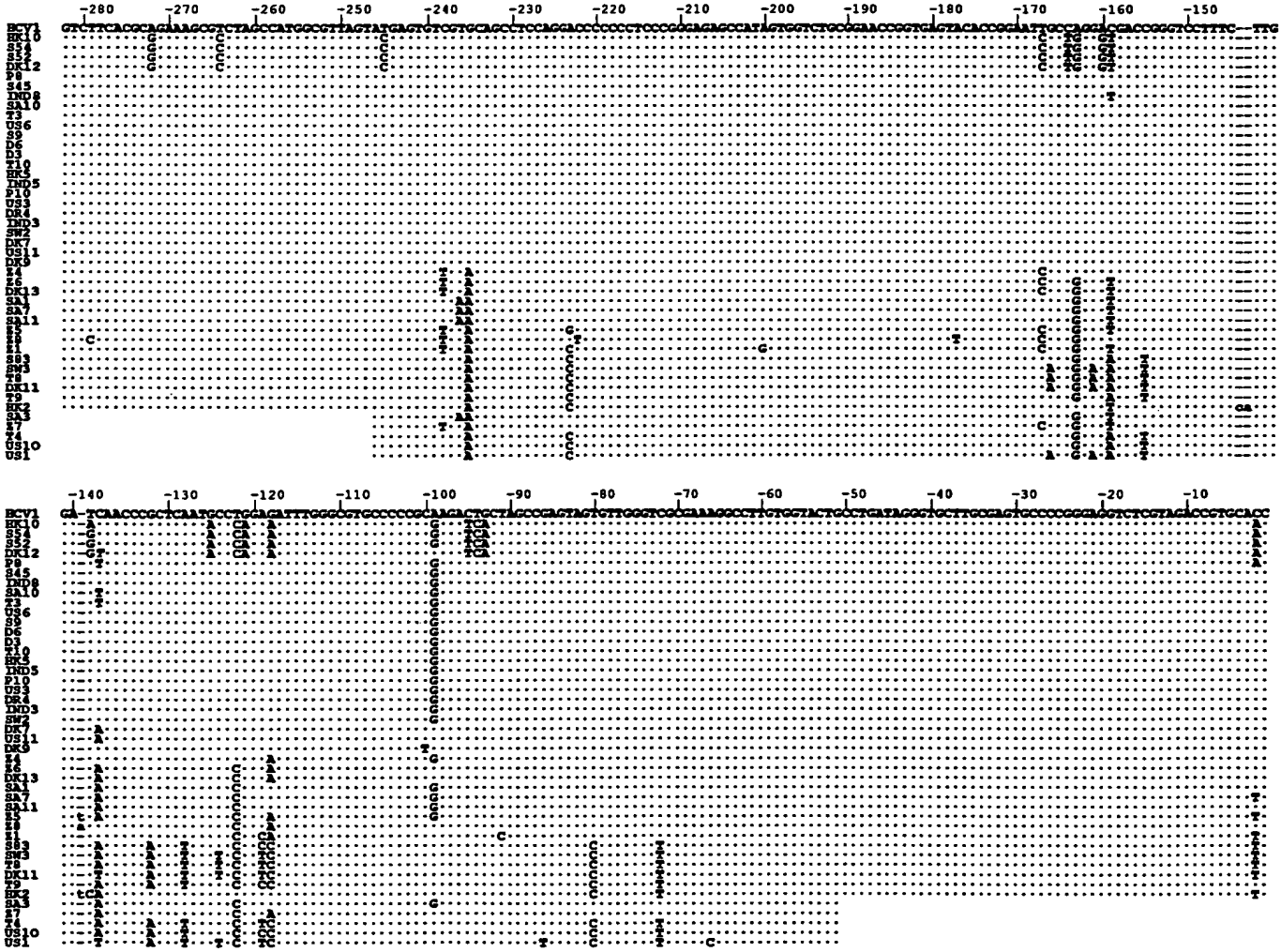


FIG. 1. Alignment of nucleotide sequences of the 5′ NC region of 44 HCV isolates from around the world. The sequences are compared to the prototype HCV sequence (HCV-1, refs. 5 and 7, and resequenced in this study) shown on the top line. Nucleotide substitutions are indicated as uppercase letters, and identical nucleotides are shown as dots. Nucleotide insertions seen in three isolates (HK2, Z5, and Z8) are shown as lowercase letters. A single site of microheterogeneity is shown in italics, at position −138 in isolate DK12 (i.e., C and T).

segregated them into three groups (HCV I, II, and III). Other published sequences can be placed in these groups (9, 10). Group I includes isolates HCV-1 (5, 7), HC-J1 (6), and HCV-H (10); group II includes isolates HC-J4 (6), HCV-J (4), HCV-BK (3), and HCV-K1 (17); and group III includes isolates HCV-K2a (17), HCV-K2b (17), and HC-J6 (9). Our analysis, based entirely on sequence of the 5′ NC region, shows that isolates from the present study represent a broad spectrum of sequence patterns that cannot all be placed within these groups. The predominant sequence pattern, seen in isolates D3, D6, DK7, DK9, DR4, HK5, IND3, IND5, IND8, P8, P10, S9, S45, SA10, SW2, T3, T10, US3, US6, and US11, was most similar to the sequence of the prototype isolate HCV-1 and closely related sequences (HCV groups I and II). In addition, there were several isolates (DK11, S83, SW3, T4, T8, T9, US1, and US10) with sequences similar to those of HCV group III. However, in the remaining isolates (DK12, DK13, HK2, HK10, S52, S54, SA1, SA3, SA7, SA11, Z1, Z4, Z5, Z6, Z7, and Z8) the 5′ NC sequence was significantly different from those of reported sequences (Fig. 1). It is noteworthy that the DK12, HK10, S52, and S54 sequences were 5–10% different from any other isolate. Thus, we have observed patterns of nucleotide sequence in the HCV 5′ NC region significantly different from the patterns assigned to groups I, II, and III (11). We were unable to segregate the sequence patterns that we observed in the 5′ NC region to defined geographical regions (Fig. 1). Further sequence analysis will demonstrate how the heterogeneity observed in the 5′ NC region among different HCV isolates relates to sequence differences elsewhere in the HCV genome.

It is well known that the primary sequence around the AUG initiation codon of a gene is important for initiation of translation (for review, see ref. 18). In this context we find it interesting that (*i*) the polyprotein start codon occurs at the same location relative to the prototype sequence in all 39 HCV isolates studied and (*ii*) the nucleotides surrounding the AUG codon are particularly well-conserved. Except for the nucleotide variation at position −2, the nucleotide sequence at positions +8 to −65 is invariant among all studied isolates. It is noteworthy that the nucleotide at position −3 contains adenine because a purine at this position is regarded as key for initiation of translation (18). Our data imply that the position of the AUG initiation codon and the surrounding sequence is crucial to the translation of the HCV polyprotein.

Different Open Reading Frames of the 5′ NC Region of the HCV Genome. Han and coworkers (7) recently described short ORFs in the 5′ NC region of the HCV and pestivirus genomes. We find that all HCV isolates examined in this study also have short ORFs. The different patterns of short ORFs observed within the 5′ NC region of the various isolates of HCV are shown in Fig. 2. We cannot comment on the presence or absence of ORF 1, described previously (5, 7), in our analysis because the initiation codon of ORF 1 in HCV-1 is 5′ to the region that we sequenced. However, most HCV isolates included in this study have three short ORFs (ORFs 2–4) identical to those described (5, 7) in the prototype HCV sequence. Interestingly, different patterns of these ORFs were seen in several HCV isolates. An initiation codon beginning at nucleotide position −160 was found in 11 isolates (DK13, HK2, IND8, SA1, SA3, SA7, SA11, Z1, Z5, Z6, and Z7). This ORF (ORF 5) was 7 amino acids in length and was fused with ORF 3 in several isolates (Fig. 2). Four isolates (DK12, HK10, S52, and S54) had only a single ORF, a version of ORF 4 longer at the 3′ end by 21 amino acids for a total of 25 amino acids. Isolate Z1 possessed the most unusual arrangement of ORFs: ORF 2 was elongated at the 3′ end by 14 amino acids for a total of 29 amino acids, and ORFs 3 and 5 were fused, creating an ORF of 49 amino acids. Thus, this isolate had ORFs spanning most of the 5′ NC region and may reflect the genomic organization of a putative ancestral virus that encoded a polyprotein extending into what is now an untranslated region. The functional status of the ORFs in the 5′ NC region of the HCV genome is unknown. We find it interesting that all 44 HCV isolates included in this study, as well as 35 HCV isolates reported by others (3–10, 16, 17, 19), have at least one ORF within the 5′ NC region. These data are consistent with the hypothesis that these ORFs are maintained because of a role in control of translation (7).

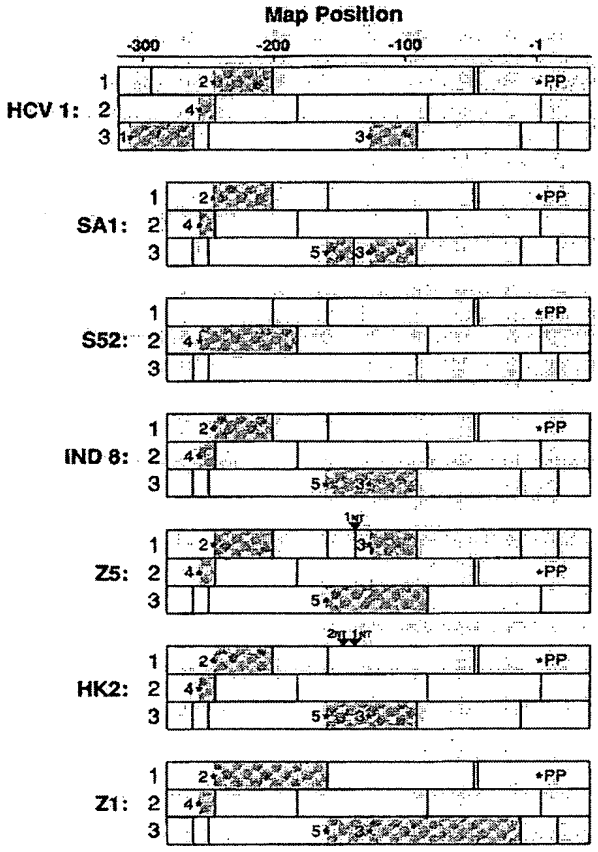


FIG. 2. Short ORFs within the 5′ NC region of the HCV genome. Shown at the top are the short ORFs described previously in the HCV prototype sequence (5, 7). All three translation frames of representative isolates are shown. The shadowed areas represent the short ORFs, defined as an initiation codon followed by codons specifying amino acids and not termination codons. Initiation codons are depicted as stars and numbered 1–5. Termination codons are depicted as vertical lines. The polyprotein start codon is marked PP. Inverted triangles show the position of nucleotide insertions. An ORF pattern similar to that of isolate SA1 was seen in isolates DK13, SA7, SA11, and Z6, and an ORF pattern similar to that of S52 was seen in isolates DK12, HK10, and S54.

Consensus Sequence of the 5′ NC Region of the HCV Genome of 81 HCV Isolates. To determine the extent of sequence variability within the 5′ NC region of the HCV genome, we combined our data on 44 HCV isolates (Fig. 1) with that of 37 reported HCV isolates (3–10, 16, 17, 19, 20) and performed a multiple sequence alignment on all HCV 5′ NC sequences currently available. The resulting consensus sequence, identical to that of the prototype sequence (5, 7), is shown in Fig. 3 as a histogram, illustrating the percent of sequences different from the consensus sequence at each nucleotide position in 282 nucleotides of the 5′ NC region of the HCV genome. Our data confirm that the 5′ NC region of the HCV genome is well conserved among HCV isolates

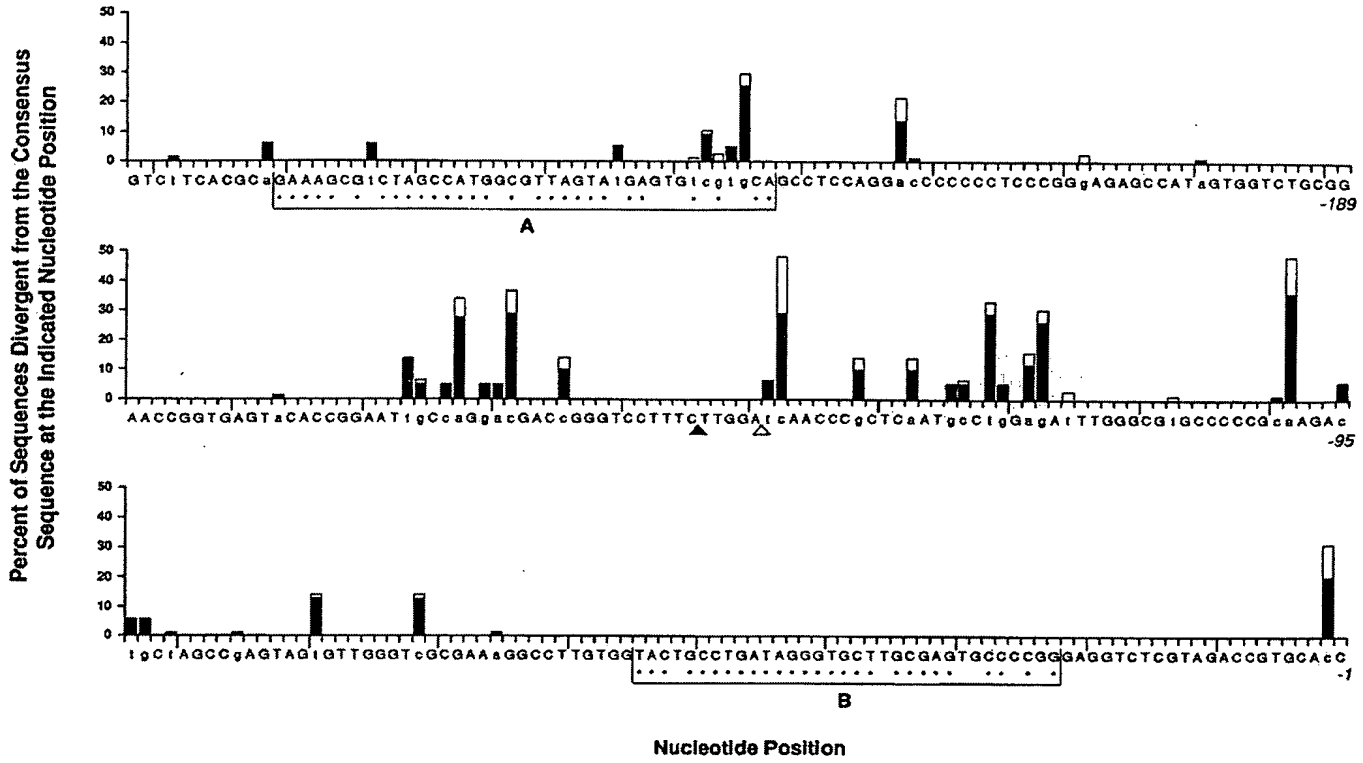


FIG. 3. Histogram of the percent of sequences different from the consensus sequence at each nucleotide position in 282 nucleotides of the 5′ noncoding region among 81 HCV isolates (37 published sequences and 44 sequences from this study) from around the world. The sequence of all 282 nucleotide positions was available in 53 isolates, of which 39 were from this study; the remaining sequences were partial sequences. The consensus sequence and the percent of sequences divergent from it are given at each nucleotide position for all sequences available at that site. Open bars represent contribution of published sequences; closed bars represent data from this study. Uppercase and lowercase letters indicate nucleotide positions that are invariant and variable, respectively. The closed triangle indicates the position of an insertion of two nucleotides seen in isolate HK2, whereas the open triangle indicates the position of insertions of a single nucleotide seen in isolates HK2, Z5, and Z8. The two domains of HCV-1 with significant similarity to bovine viral diarrhea virus (BVDV) are boxed (spaces between adjacent nucleotides required for optimal alignment of HCV-1 and BVDV are omitted in this figure). Dots under a nucleotide within boxes A and B indicate identical nucleotide matches between HCV-1 and BVDV.

because the overall nucleotide variation, defined as the total number of nucleotides different from the consensus sequence, is only 2.0%. However, we find that the 5′ NC region of HCV consists of highly conserved domains interspersed between variable domains. Nucleotide variations from the consensus sequence are found in 45 (16%) of the 282 nucleotide positions. The most variable domain spans 50 nucleotides (positions −167 to −118) and has an overall nucleotide variation of 6.0% with variation from the consensus sequence at 18 (36%) of the nucleotide positions. In addition, the nucleotide insertions observed in three HCV isolates in this study are located within this domain (Fig. 1). It is noteworthy that the nucleotide identity in this variable domain between two different HCV isolates (HK10, S52, S54 versus DK11, SW3, T8, US1, and DK12 versus SW3, T8) is as low as 68%, and the identity to HCV-1 is as low as 76% (DK11, SW3, T8, US1) (Fig. 1). Furthermore, this variable domain has two subdomains with even greater variability flanking a region that is invariant, except for the nucleotide insertions observed in three isolates. The overall nucleotide variability of these subdomains (positions −167 to −155 and −139 to −118) is 9.3% and 8.1%, respectively, and nucleotide variations from the consensus sequence were seen at 8 (61.5%) and 10 (45.5%) of the nucleotide positions, respectively. We have thus defined a variable domain of 50 nucleotides within the 5′ NC region of HCV with significant heterogeneity among different HCV isolates. Two other variable domains could be identified in the HCV 5′ NC region (positions −239 to −222 and −100 to −72) that displayed an overall nucleotide variation of 4.0% and 3.4%, respectively. Nucleotide variations from the consensus sequence were seen at 7 (38.9%) and 9 (31%) of these nucleotide positions, respectively. The high degree of nucleotide changes seen within the variable domains suggests that functional constraints are low in these regions.

The remaining 185 of the 282 nucleotides of the HCV 5′ NC region analyzed are highly conserved with an overall nucleotide variation of only 0.3%. Nucleotide variations from the consensus sequence were seen at only 11 (5.9%) of the 185 nucleotide positions. Interestingly, there are three long stretches with completely invariant nucleotide sequences of 18, 22, and 63 bases (positions −263 to −246, −199 to −178, and −65 to −3, respectively) among all studied HCV isolates. Most impressive is the stretch of 63 invariant nucleotides immediately upstream of the polyprotein start codon. This region contains a domain with significant similarity to pestiviruses (see below). The 5′ NC region of a viral genome typically contains regulatory elements, so it is likely that this region in HCV is conserved because it contains cis-acting elements involved in replication of the viral genome (e.g., RNA packaging signal, etc.) or expression of viral genes (e.g., translation initiation signal, etc.).

Previously reported sequence analysis suggests a distant evolutionary relationship among certain proteins of HCV, flaviviruses, and pestiviruses (15, 21). We used computer-assisted nucleotide sequence analysis to look for similarity

within the 5′ NC regions of these viruses. Using the program SEQ, we found no such similarity between HCV and flaviviruses. However, two nucleotide domains within the 5′ NC region of the HCV-1 genome showed statistically significant similarity to 5′ NC sequences of pestiviruses. (*i*) A domain in the 5′ end of the 5′ NC region (box A in Fig. 3) of HCV-1 was found to have statistically significant similarity ($P = 0.003$) with BVDV (22) but not with hog cholera virus (23, 24). (*ii*) Next, a domain in the 3′ end of the 5′ NC region (box B in Fig. 3) of HCV-1 was found to have significant similarity ($P = 0.001$) with BVDV (22) and two strains of hog cholera virus ($P = 0.002$, ref. 23 and $P = 0.005$, ref. 24). These two nucleotide domains correspond to conserved regions I and IV described by Han and coworkers (7) in an alignment of the 5′ NC region of HCV and pestiviruses. The two domains within the 5′ NC region of HCV with significant similarity to pestiviruses are likely to have been conserved in virus evolution because of an important biological role. This hypothesis is supported by our finding that these domains are conserved among many different HCV isolates (Fig. 3).

The findings in this study have important implications for the selection of primers in cDNA PCR assays to detect HCV RNA because genetic heterogeneity among different HCV strains results in false negative results because of primer and template mismatch. The 5′ NC region, previously shown to be the most conserved region of the HCV genome, was a natural first choice for designing primers for cDNA PCR assays (for detailed review, see ref. 11). However, in this comprehensive analysis, we have identified additional variable sequences within the 5′ NC region of the HCV genome that should be avoided in the design of primers for cDNA PCR assays. Conversely, our data highlight regions within the 5′ NC region of the HCV genome that are especially well conserved among a large number of different HCV isolates from around the world. The long stretches of invariant nucleotides are preferred for primer design. We have recently shown in a study comparing primers that a cDNA PCR assay with primers designed from the two domains of the HCV 5′ NC region that share statistically significant similarity with BVDV sequences (i.e., boxes A and B in Fig. 3) is both sensitive and specific for detecting HCV RNA (12). We have now further demonstrated that these domains are highly conserved in a large number of HCV isolates from around the world.

In summary, we have demonstrated significant nucleotide sequence variation within the 5′ NC region of HCV in several HCV isolates. Furthermore, we have defined highly conserved domains within the 5′ NC region of HCV, which suggest that these domains have crucial functional roles.

We thank Ms. T. Tsareva for synthesis of oligonucleotides and Mr. T. J. Heishman for assistance with graphics and manuscript preparation. We gratefully acknowledge the generosity of the following investigators for providing sera for this study: Drs. H. Alter, V. Arankalle, D. W. Bradley, D.-S. Chen, P. Farci, M. C. Kew, K. Krogsgaard, A. Lok, T. Quinn, F. Renger, M. Sjogren, and A. Widell. Computer-assisted nucleic acid analysis was through the GenBank Online Service.

1. Choo, Q.-L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W. & Houghton, M. (1989) *Science* **244**, 359–362.
2. Kuo, G., Choo, Q.-L., Alter, H. J., Gitnick, G. L., Redeker, A. G., Purcell, R. H., Miyamura, T., Dienstag, J. L., Alter, M. J., Stevens, C. E., Tegtmeier, G. E., Bonino, F., Colombo, M., Lee, W.-S., Kuo, C., Berger, K., Shuster, J. R., Overby, L. R., Bradley, D. W. & Houghton, M. (1989) *Science* **244**, 362–364.
3. Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onishi, E., Andoh, T., Yoshida, I. & Okayama, H. (1991) *J. Virol.* **65**, 1105–1113.
4. Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T. & Shimotohno, K. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9524–9528.
5. Choo, Q.-L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, A., Barr, P. J., Weiner, A. J., Bradley, D. W., Kuo, G. & Houghton, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2451–2455.
6. Okamoto, H., Okada, S., Sugiyama, Y., Yotsumoto, S., Tanaka, T., Yoshizawa, H., Tsuda, F., Miyakawa, Y. & Mayumi, M. (1990) *Jpn. J. Exp. Med.* **60**, 167–177.
7. Han, J. H., Shyamala, V., Richman, K. H., Brauer, M. J., Irvine, B., Urdea, M. S., Tekamp-Olson, P., Kuo, G., Choo, Q.-L. & Houghton, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1711–1715.
8. Cha, T.-A., Kolberg, J., Irvine, B., Stempien, M., Beall, E., Yano, M., Choo, Q.-L., Houghton, M., Kuo, G., Han, J. H. & Urdea, M. S. (1991) *J. Clin. Microbiol.* **29**, 2528–2534.
9. Okamoto, H., Okada, S., Sugiyama, Y., Kurai, K., Iizuka, H., Machida, A., Miyakawa, Y. & Mayumi, M. (1991) *J. Gen. Virol.* **72**, 2697–2704.
10. Inchauspe, G., Zebedee, S., Lee, D.-H., Sugitani, M., Nasoff, M. & Prince, A. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10292–10296.
11. Houghton, M., Weiner, A., Han, J., Kuo, G. & Choo, Q.-L. (1991) *Hepatology* **14**, 381–388.
12. Bukh, J., Purcell, R. H. & Miller, R. H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 187–191.
13. Bachmann, B., Luke, W. & Hunsmann, G. (1990) *Nucleic Acids Res.* **18**, 1309.
14. Winship, P. R. (1989) *Nucleic Acids Res.* **17**, 1266.
15. Miller, R. H. & Purcell, R. H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2057–2061.
16. Ogata, N., Alter, H. J., Miller, R. H. & Purcell, R. H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 3392–3396.
17. Nakao, T., Enomoto, N., Takada, N., Takada, A. & Date, T. (1991) *J. Gen. Virol.* **72**, 2105–2112.
18. Kozak, M. (1991) *J. Cell Biol.* **115**, 887–903.
19. Fuchs, K., Motz, M., Schreier, E., Zachoval, R., Deinhardt, F. & Roggendorf, M. (1991) *Gene* **103**, 163–169.
20. Takeuchi, K., Kubo, Y., Boonmar, S., Watanabe, Y., Katayama, T., Choo, Q.-L., Kuo, G., Houghton, M., Saito, I. & Miyamura, T. (1990) *J. Gen. Virol.* **71**, 3027–3033.
21. Choo, Q.-L., Han, J., Weiner, A. J., Overby, L. R., Bradley, D. W., Kuo, G. & Houghton, M. (1991) in *Viral Hepatitis C, D, and E*, eds. Shikata, T., Purcell, R. H. & Uchida, T. (Elsevier, Amsterdam), pp. 47–52.
22. Collett, M. S., Larson, R., Gold, C., Strick, D., Anderson, D. K. & Purchio, A. F. (1988) *Virology* **165**, 191–199.
23. Meyers, G., Rumenapf, T. & Thiel, H.-J. (1989) *Virology* **171**, 555–567.
24. Moormann, R. J. M., Warmerdam, P. A. M., van der Meer, B., Schaaper, W. M. M., Wensvoort, G. & Hulst, M. M. (1990) *Virology* **177**, 184–198.